Environmental Sciences Europe

Open Access

Establish data infrastructure to compile and exchange environmental screening data on a European scale

Jaroslav Slobodnik[1], Juliane Hollender[2], Tobias Schulze[3], Emma L. Schymanski[4] and Werner Brack[3,5*]

Abstract

Robust techniques based on liquid (LC) and gas chromatography (GC) coupled with high-resolution mass spectrometry (HR-MS) enable sensitive screening, identification, and (semi)quantification of thousands of substances in a single sample. Recent progress in computational sciences has enabled archiving and processing of HR-MS 'big data' at the routine level. As a result, community-based databases containing thousands of environmental pollutants are rapidly growing and large databases of substances with unique identifiers allowing for inter-comparison at the global scale have become available. A data-archiving infrastructure is proposed, allowing for retrospective screening of HR-MS data, which will help define the 'chemical universe' of organic substances and enable prioritisation of toxicants causing adverse environmental effects at the local, river basin, and national and European scale in support of the European water and chemicals management policy.

Challenge

Non-target screening (NTS) workflows are a powerful method for the large-scale analysis of environmental samples. They consist of wide-scope target, suspect, and non-target analysis. Recently, NTS has developed rapidly with the advance of HR-MS techniques, as reviewed elsewhere [1]. Smart monitoring combining cost-effective methods for wide-scope target and suspect screening with a battery of well-established high-throughput bioassays could be used routinely to reduce the risk of overlooking toxic chemicals in the environment [2, 3].

Continental scale wide-scope target and non-target screening required for an appropriate monitoring of complex chemical contamination is rapidly developing in many monitoring laboratories, as recommended in [4]. This will provide an amount of information unprecedented so far in environmental monitoring. Currently, monitoring data are typically stored and evaluated in a closed and decentralised way using non-harmonised

formats and without substantial data exchange between the scientists and agencies involved. These deficiencies hamper the recognition of newly emerging contaminants and mixtures, the prioritisation and identification of the newly recognised chemicals, and the efficient exploitation of these data for quality assessment and management on a European and even global scale. So far, the infrastructure for storage, long-term archiving, open exchange, processing and analysis of these data is largely lacking, although the required technology for 'big data' repositories is already available [1, 5].

Any LC-HR-MS or GC-HR-MS technique needed for the detection of suspect and non-target chemicals generates large amounts of data, up to tens of GB per analysis. This brings environmental monitoring into the arena of 'big data'. Currently, only a fraction of the information from HR-MS measurements is extracted and the rest is discarded. The challenge is (i) to extract the minimum necessary information for a quick overview of presence/absence of a large number of suspects in the samples and (ii) to save all information from HR-MS (raw data) in a format harmonised at the European (and possibly global) level for retrospective screening of environmental samples for the currently known and future pollutants.

*Correspondence: werner.brack@ufz.de
[3] UFZ Helmholtz Centre for Environmental Research GmbH, Permoserstraße 15, 04318 Leipzig, Germany
Full list of author information is available at the end of the article

Dealing with tens of thousands of substances, their transformation products, technical mixtures, salts, isomers, etc. may lead to a great confusion when not coordinated. Neither the CAS No. nor the name is a sufficiently unique identifier for a compound of interest. At present, the US EPA CompTox Chemicals Dashboard (https://comptox.epa.gov/dashboard; >875,000 chemicals, [6]) is used as a reference for extracting quality checked information. Still, many of the chemicals with high production volumes and their transformation products are not found in this or any databases.

The identification of compounds with experimentally obtained mass spectra is more reliable than just exact mass matching of compound databases [7]. To ensure this, community-based databases containing measured mass spectra need to grow considerably. In addition, the mass spectra of 'unknowns' frequently recorded in environmental samples should be stored for future identification, as done in prototype form in the European (NORMAN) MassBank (https://massbank.eu/MassBank/).

Complex mixtures of chemicals should be considered together with their complex effects and ecosystem impacts. Technical developments that now allow for recording extensive chemical fingerprints from NTS, toxicity profiles, and omics responses in laboratory test systems and wildlife and environmental DNA to address biodiversity are delivering enormous amounts of data. The challenge is to establish the infrastructure needed for data storage and the tools for multivariate biological and chemical analysis to facilitate the use of such data.

Recommendations

- Establish a federated European infrastructure storing raw non-target screening data converted into a common (open) format allowing for 'on demand' accessibility for retrospective screening
- Establish a central platform/database storing regularly updated information on available data sets Europe-wide and, eventually, at a global scale
- Establish a common European platform where the unique identifiers of newly discovered environmental pollutants can be shared in a harmonised format
- Apply commonly agreed workflow(s) for retrospective analysis to identify and prioritise pollutants frequently detected in environmental samples.

Requirements

Establishing the data infrastructure for compilation and exchange of screening data on a European scale requires:

- Recognising the need for screening data within the framework of European water policy, air and soil pollution, and waste management
- Providing incentives by the European Commission to scientists, monitoring agencies, and Member States to share the screening data
- Providing incentives by the scientific journals to scientists to share the raw screening data in a harmonised format as a supplementary information to the publications using these data
- Securing European and national scale funding for establishment of the interoperable infrastructure
- Support of the European MassBank for systematic storage of mass spectral information of environmentally relevant substances (https://massbank.eu/MassBank)
- Further harmonisation of wide-scope target and suspect screening techniques in Europe
- Further development of HR-MS data processing workflows.

Achievements

SOLUTIONS/NORMAN database system

The NORMAN network (https://www.norman-network.net); a network of more than 80 reference laboratories, research centres and other organisations for monitoring of emerging environmental substances in Europe and North America; [8]) and the SOLUTIONS project (https://www.solutions-project.eu); [9]) have pushed the limits of NTS further using European case studies. It is now possible to screen more than 2000 target compounds and more than 40,000 suspect substances in environmental samples. An online database for wide-scope target and non-target screening data was developed as a part of the NORMAN Database System (https://www.norman-network.com/nds) and the SOLUTIONS Database System (https://www.norman-network.com/solutions/norman.php). The latter contains also a unique list of modelling-based prioritised substances, whose presence in the environment is not determined on actual occurrence measurements, but rather on the predictions related to their production volumes, use pattern, and how easy they can be released into environment.

NORMAN suspect list exchange

A collaborative trial organised by the NORMAN network on a surface water sample from the Danube river basin revealed that suspect screening using specific lists of chemicals to find "known unknowns" was a very common and efficient way to expedite non-target screening [10]. As a result, the NORMAN Suspect List Exchange was founded (https://www.norman-network.com/nds/

Slobodnik *et al. Environ Sci Eur*   (2019) 31:65

Page 3 of 4

SLE/) and members were encouraged to submit their suspect lists. To date, more than 50 lists of highly varying substance numbers have been uploaded. Over 40,000 substances are available in the correspondingly merged SusDat database (https://www.norman-network.com/nds/susdat). This database contains harmonised names, CAS Nos., SMILES, InChIKeys, "MS-ready structure forms" with chemical substances provided in the form observed by the mass spectrometer (e.g., desalted, as separate components of mixtures [11]), exact masses, retention indices, and modelling-based predicted ecotoxicity threshold values. Further > 40,000 substances are in the pipeline. The curation was done within the network using open-access cheminformatics toolkits. Starting in 2017, the NORMAN Suspect List Exchange and US EPA CompTox Chemicals Dashboard (https://comptox.epa.gov/dashboard) pooled resources in curating and uploading these lists to the Dashboard (https://comptox.epa.gov/dashboard/chemical_lists).

NORMAN digital sample freezing platform (DSFP)

A retrospective screening platform for hosting mass spectrometric data obtained by LC-HR-MS was created in 2017 (https://norman-data.net), with the ambition of becoming a European and possibly global standard for retrospective suspect screening of environmental pollutants [5; Fig. 1]. This platform enables a quick and effective overview of the potential presence of thousands of substances either known or suspected to be present in the environment (based on the SusDat database), including a wide range of contaminants of emerging concern, their transformation products and unknowns, across a large number of samples and different matrices. A tool for semi-quantitative estimation of concentrations of any detected compound based on their structure similarity is being tested.

European (NORMAN) MassBank

A database for MS (mainly high resolution) spectra of substances of environmental and metabolomic relevance was created in Europe in 2011, using a format developed previously in Japan. European (NORMAN) MassBank (https://massbank.eu/MassBank/) now contains 57,472 unique mass spectra of 14,667 substances (accessed on 10 May 2019). The exact mass, fragmentation, and measurement information on all substances are feeding into the NORMAN DSFP. In SOLUTIONS, the joint efforts of the environmental and metabolomics community on MassBank development improved and a developer consortium was founded (https://github.com/MassBank/).

Demonstration and evaluation in case studies

The databases developed within NORMAN/SOLUTIONS presented above have already been applied in

Fig. 1 Adopted workflow for obtaining harmonised raw screening monitoring data through the Digital Sample Freezing Platform (DSFP) interface [5]

Slobodnik *et al. Environ Sci Eur*   (2019) 31:65

Page 4 of 4

several case studies related to SOLUTIONS. In the Joint Danube Survey 3 (2013; [12]), a wide-scope target and suspect screening using comprehensive substance lists was tested by several laboratories. Wide-scope target screening tools combined with bioassays were systematically used at the assessment of abatement options in the River Rhine catchment [13]. The NormaNEWS study was carried out in 2017, establishing a global emerging contaminant early warning network to rapidly assess the spatial and temporal distribution of contaminants of emerging concern in environmental samples through performing retrospective analysis on HR-MS data. The effectiveness of such a network was demonstrated through a pilot study, in which eight reference laboratories with available archived HR-MS data retrospectively screened data acquired from aqueous environmental samples collected in 14 countries on 3 different continents [14]. Wide-scope target (> 2100 substances) and suspect screening (NORMAN SusDat; > 40,000 substances) were performed in water, sediment, and biota samples in the Joint Black Sea Surveys (2016, 2017; [15]). A thorough analysis of waste water treatment plant effluents with a battery of SOLUTIONS/NORMAN bioassays was applied using wide-scope target and suspect screening in the Danube River Basin in 2017 in cooperation with the International Commission for the Protection of the Danube River (ICPDR) [16]. The outcomes of the case studies support further development of harmonised databases for archiving 'big data' from NTS.

Authors' contributions
JS and WB conceptualized and drafted the manuscript. The other authors elaborated the manuscript and contributed specific aspects. All authors read and approved the final manuscript.

Availability of data and materials
Not applicable; presented information is based on previously published data only.

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1] Environmental Institute, Okruzna 784/42, 97241 Kos, Slovak Republic. [2] Eawag, Swiss Federal Institute of Aquatic Science and Technology, Überlandstrasse 133, 8600 Dübendorf, Switzerland. [3] UFZ Helmholtz Centre for Environmental Research GmbH, Permoserstraße 15, 04318 Leipzig, Germany. [4] Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, 6, Avenue du Swing, 4367 Belvaux, Luxembourg. [5] Department of Ecosystem Analysis, Institute for Environmental Research, ABBt-Aachen Biology, Aachen, Germany.

References
1. Hollender J et al (2017) Nontarget screening with high resolution mass spectrometry in the environment: ready to go? Environ Sci Technol 51:11505–11512
2. Brack W et al (2018) Towards a holistic and solution-oriented monitoring of chemical status of European water bodies: how to support the EU strategy for a non-toxic environment? Environ Sci Eur 30:33
3. Altenburger R et al (2015) Future water quality monitoring—adapting tools to deal with mixtures of pollutants in water resource management. Sci Total Environ 512:540–551
4. Brack W et al (2019) High-resolution mass spectrometry to complement monitoring and track emerging chemicals and pollution trends in European water resources. Environ Sci Eur. https://doi.org/10.1186/s12302-019-0230-0
5. Alygizakis N et al (2019) NORMAN Digital Sample Freezing Platform; A European virtual platform to exchange liquid chromatography high resolution-mass spectrometry data and screen suspects in "digitally frozen" environmental samples. Trends Anal Chem 115:129–137. https://doi.org/10.1016/j.trac.2019.04.008
6. Williams et al (2017) The CompTox Chemistry Dashboard: a community data resource for environmental chemistry. J Cheminform 9:61
7. Schymanski et al (2014) Identifying small molecules via high resolution mass spectrometry: communicating confidence. Environ Sci Technol 48(4):2097–2098
8. Dulio V et al (2018) Emerging pollutants in the EU: 10 years of NORMAN in support of environmental policies and regulations. Environ Sci Eur 30:5
9. Brack W et al (2015) The SOLUTIONS project: challenges and responses for present and future emerging pollutants in land and water resources management. Sci Total Environ 503(3):22–31
10. Schymanski E et al (2015) Non-target screening with high-resolution mass spectrometry: critical review using a collaborative trial on water analysis. Anal Bioanal Chem 407:6237–6255
11. McEachran AD et al (2018) "MS-Ready" structures for non-targeted high-resolution mass spectrometry screening studies. J Cheminform. https://doi.org/10.1186/s13321-018-0299-2
12. Liska I et al. Joint Danube Survey 3: a comprehensive analysis of danube water quality. http://www.danubesurvey.org/jds3/jds3-files/nodes/documents/jds3_final_scientific_report_1.pdf. 2015. ISBN: 978-3-200-03795-3
13. Neale PA et al (2017) Integrating chemical analysis and bioanalysis to evaluate the contribution of wastewater effluent on the micropollutant burden in small streams. Sci Total Environ 576:785–795
14. Alygizakis N et al (2018) Exploring the potential of a global emerging contaminant early warning network through the use of retrospective suspect screening with high-resolution mass spectrometry. Environ Sci Technol 52(9):5135–5144
15. Slobodnik et al. (2016) National Pilot Monitoring Studies and Joint Open Sea Surveys in Georgia, Russian Federation and Ukraine. http://emblasproject.org/wp-content/uploads/2018/08/EMBLAS-II_NPMS_JOSS_2016_ScReport_Final3.pdf
16. Alygizakis et al (2019) Characterization of wastewater effluents in the Danube River Basin with chemical screening, in vitro bioassays and antibiotic resistant genes analysis. Environ Int 127:420–429. https://doi.org/10.1016/j.envint.2019.03.060

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.